RNA-DNA Triplex Formation by Long Noncoding RNAs

AUTHOR(S):

Li, Yue; Syed, Junetha; Sugiyama, Hiroshi

CITATION:

Li, Yue ...[et al]. RNA-DNA Triplex Formation by Long Noncoding RNAs. Cell Chemical Biology 2016, 23(11): 1325-1333

ISSUE DATE:

2016-11-17

URL:

http://hdl.handle.net/2433/230817

RNA–DNA triplex formation by long noncoding RNAs

Yue Li, [1] Junetha Syed, [1] and Hiroshi Sugiyama [1, 2, *]

[1]Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa-oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan
[2]Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Yoshida-ushinomiyacho, Sakyo-ku, Kyoto 606-8501, Japan
*Correspondence: hs@kuchem.kyoto-u.ac.jp

Abstract

Long noncoding RNAs (lncRNAs) play a pivotal role in the regulation of biological processes through various mechanisms that are not fully understood. Proposed mechanisms include regulation based on RNA–protein interactions, as well as RNA–RNA interactions and RNA–DNA interactions. Here, we focus on one possible mechanism that lncRNA might be using to impact biological function, the RNA–DNA triplex formation. We summarize currently available examples of lncRNA triplex formation and discuss the details surrounding orientation of triplex formation as one of the key properties guiding this process. We propose that symmetrical triplex-forming motifs, especially those in *cis*-acting lncRNAs, favor triplex formation. We also consider the effects of lncRNA structures, protein or ligand binding, and chromatin structures on the lncRNAs triplex formation.

Introduction

Over the past decade, advances in the field of genome sequencing technologies have provided a deeper understanding of the complex genomic architecture (Kapranov et al., 2007; Consortium, 2004, 2012). The most surprising insights that genome sequencing has revealed is that only a small fraction of the mammalian DNA encodes protein-coding messenger RNA (mRNA). However, that does not mean that the rest of the DNA is not transcribed, and we now appreciate that a large proportion of transcribed RNA corresponds to noncoding RNAs (ncRNAs), which are thought to play roles in important biological processes, particularly epigenetic regulation (Wilusz et al., 2009; Ponting et al., 2009; Guttman et al., 2009; Kung et al., 2013; Morris and Mattick, 2014).

Over the last several years, we have seen a steady increase in the interest in identifying ncRNAs and understanding their regulatory functions (Quinn and Chang, 2016; Bonasio and Shiekhattar, 2014; Batista and Chang, 2013; Lee, 2012). ncRNAs include a wide class of regulatory RNAs that can be classified into small ncRNAs such as microRNAs (miRNAs), small interfering RNAs (siRNAs), PIWI-interacting RNAs (piRNAs), and long noncoding RNAs (lncRNAs), which are more than 200 nt in length (Cech and Steitz, 2014; Kowalczyk et al., 2012). Majority of mammalian ncRNAs, including lncRNAs, are highly conserved, strongly supporting the view that these RNAs play important *in vivo* roles (Struhl, 2007). Several recent studies have provided evidence for biological functions of lncRNAs, especially in the regulation of gene expression at various levels (Rinn and Chang, 2012; Mattick, 2004). These findings are consistent with the observation that a large number of lncRNAs are transcribed from loci with proximity to protein-coding genes, such as enhancer RNAs (eRNAs) and upstream antisense RNAs (uaRNAs) (Bonasio and Shiekhattar, 2014).

lncRNAs regulate gene expression through complex mechanisms that are not fully understood (Guttman and Rinn, 2012; Hung and Chang, 2010). It has been suggested that lncRNAs fulfill their regulatory functions by serving as molecular signals, decoys, guides, and scaffolds (Wang and Chang, 2011). The molecular mechanisms of lncRNAs can be simplified into these four main archetypes, but an individual lncRNA is not confined to one archetype and may function using several different mechanisms. For example, during X chromosome inactivation (XCI), *Xist* RNA acts as both a molecular signal for the chromosome-wide repression of gene expression and a guide for the recruitment of polycomb repressive complex 2 (PRC2) (Van Bemmel et al., 2016; Maclary et al., 2013; Galupa and Heard, 2015; Zhao et al., 2008). Telomeric repeat-containing RNA (TERRA) can bind to the human telomere-binding protein TLS/FUS, recruit histone-modifying enzymes, and regulate telomere length through histone modifications (Takahama and Oyoshi, 2013; Takahama et

al., 2013). As these two examples suggest, lncRNA biological mechanisms converge on establishing intermolecular interactions with other biomolecules, such as proteins, DNA and other RNAs. Therefore, one practical approach to expand our understanding of the lncRNA molecular mechanisms is to elucidate the RNA–protein and RNA–DNA interactions they engage in. lncRNAs can form RNA–protein complexes, which most commonly mediate the correct localization of chromatin regulatory proteins (*Xist* RNA) or recruitment of functional proteins with distinct protein interaction domains (*HOTAIR*) (Chu et al., 2015; Sarma et al., 2014; Tsai et al., 2010). In addition, lncRNAs can target specific DNA sequences by forming structures such as RNA–DNA hybrid duplexes or the RNA–DNA triplexes (Wang and Chang, 2011; Li et al., 2016). Given the structural complexity of lncRNAs, several different interactions might be involved in the molecular mechanisms relating to each individual lncRNA.

Several excellent reviews have covered the biological roles of RNA-DNA triplexes in general (Duca et al., 2008; Bacolla et al., 2015; Buske et al., 2011; Jain et al., 2008) and the molecular mechanisms of lncRNAs (Guttman and Rinn, 2012; Hung and Chang, 2010; Wang and Chang, 2011). Here, we focus on the current knowledge about RNA–DNA triplex formation by lncRNAs. We start by discussing the structure of RNA–DNA triplexes and summarizing recently reported examples of RNA–DNA triplex formation by lncRNAs. We then discuss the importance of the orientation of triplex formation to the mechanism of lncRNA actions and the possible ways proteins and ligands may regulate triplex formation.

Discovery and nature of triplex structures

Not long after the discovery of the DNA double-helix structure in 1953 (Watson and Crick, 1953), Felsenfeld et al. suggested the possibility of triplex formation, whereby two pyrimidine strands and one purine strand would form a more complex structure (Felsenfeld et al., 1957). Over the decades, the evidence of triplex formation went from a scientific curiosity (Morgan and Wells, 1968; Beal and Dervan, 1991), to inspiring researchers to apply the concept of the triple helix to develop various sequence-specific triplex-forming oligonucleotides (TFOs) as DNA cleavage tools with major biological applications in chromosome mapping (Moser and Dervan, 1987; Le Doan et al., 1987; Broitman et al., 1987; Wells et al., 1988).

In the triplex structure, a third strand can insert into the major groove of the duplex structure with sequence specificity (Hoogsteen, 1959; De los Santos et al., 1989; Koshlap et al., 1997; Pasternack et al., 2002; Radhakrishnan and Patel, 1994a, 1994b). As shown in Figure 1, each triplet can be formed by either Hoogsteen hydrogen bonding or reverse Hoogsteen hydrogen bonding between the third

base and the Watson–Crick base pair. Two types of triplexes can be formed based on the orientation of the third strand. Parallel triplex structures are formed by TA T, CG·C^+, and CG·G triplets, and antiparallel triplexes are formed by TA·A, TA·T and CG·G. In the case of CG·C^+, an acidic condition is favorable for triplex formation because the protonation of cytosine is required for effective triplet formation. Using these principles, one can determine the orientation of the third strand from its sequence (Figure 1B). The third strand with T(U)C motifs prefers to form parallel triplexes, whereas AG motifs prefer antiparallel triplexes, and GT(U) motifs can form either parallel or antiparallel triplexes.

The Hoogsteen hydrogen bonding by the third strand is weaker than the Watson–Crick hydrogen bonding in the duplex structure. Therefore, there are some limitations to the formation of the triplex structure, especially under physiological conditions (Duca et al., 2008). As mentioned above, an acidic condition is required for the protonation of cytosine during the formation of the CG·C^+ triplet (De los Santos et al., 1989). Multivalent cations such as Mg^{2+} are frequently used *in vitro* to neutralize the unfavorable charge repulsion among three negatively charged nucleic acid strands (Floris et al., 1999). Molecular crowding conditions that mimic the cell environment are suggested to increase the melting temperature of triplex structures by the hydration effect (Nakano et al., 2014; Miyoshi and Sugimoto, 2008). Another study has shown that molecular crowding induces the formation of the i-motif structure by triplet repeat DNA oligomers under neutral pH, which implies an effect similar to the protonation of cytosine in triplex formation (Rajendran et al., 2010). To overcome these limitations, chemical modifications, including base, sugar, and backbone modifications, are used widely in the design of TFOs (Torigoe et al., 2001; Lacoste et al., 1997; Hojland et al., 2007; Roberts and Crothers, 1992; Lacroix et al., 2000). These modifications may facilitate triplex formation by increasing Hoogsteen hydrogen bonding or by restricting the third strand at the desired conformation by blocking the secondary structure formation, or other mechanisms.

Given that triplex formation requires base pairing that is weaker than Watson-Crick or unusual protonation states for the bases involved, it is valid and an important question to ask whether triplex structures exist *in vivo*. Many bioinformatics tools have been developed to analyze the potential for the presence of triplex-forming motifs in the genome (Goñi et al., 2004; Buske et al., 2012). Bioinformatics analyses have suggested that a large population of triplex-forming motifs is present across the genome. Interestingly, these motifs tend to accumulate in the gene-regulatory regions, particularly in promoter regions, leading to a proposal that triplex formation might play significant regulatory roles *in vivo*. More direct evidence of triplex formation *in vivo* comes from using triplex-binding antibodies, as well as fluorescent molecules that selectively recognize triplex

structures (Lee et al., 1987; Agazie et al., 1994, 1996; Lubitz et al., 2010). The results of studies using immunofluorescence of triplex-specific antibodies strongly suggest the formation of the triplex structure in the nucleus. Most importantly, these antibodies exhibit a stronger affinity for RNA–DNA triplexes than for DNA triplexes, which indicates the formation of RNA–DNA triplexes *in vivo*. Recent studies have also demonstrated the presence of specific proteins that can interact with triplexes *in vivo*. For instance, helicases such as RecQ have been shown to unwind triplex structures, which strongly suggests a potential regulatory role of triplexes *in vivo* (Jain et al., 2010; Maine and Kodadek, 1994; Dixon et al., 2008; Brosh et al., 2001).

Consistent with the RNA–DNA triplex, studies have suggested that RNA, as the third strand, forms more stable triplexes than its DNA counterparts (Roberts and Crothers, 1992). The first example of triplex formation between RNA polymers and the formation of intramolecular triplexes of RNA are well characterized. A 2′-methoxy group has also been introduced to stabilize the C3′-endo conformation of the sugar and to promote RNA-mediated triplex formation (Shimizu et al., 1992; Asensio et al., 1999). Therefore, third-strand RNA is favorable for the formation of the triplex. Recent studies have provided examples of lncRNA-mediated RNA–DNA triplex structures. In the following section, we discuss the reported lncRNAs that have been shown to form triplex structures.

lncRNA:DNA:DNA triplexes *in vivo*

Despite being recognized as a structure that RNA and DNA molecules can form *in vitro* very early on in the history of structural biology of nucleic acids, it took another 50 years to begin describing triplex structures *in vivo* and documenting their physiological relevance. In 2007 Martianov et al. reported triplex formation by a noncoding RNA that was transcribed as an interfering transcript from the minor promoter of the human gene encoding dihydrofolate reductase (*DHFR*) (Martianov et al., 2007). This interfering RNA transcript had a repressive effect on the transcription of *DHFR* mRNA from the major *DHFR* promoter located downstream. This effect was observed only when the sequence of interfering transcript overlapped with the sequence of the core major *DHFR* promoter. Subsequent studies have shown that the noncoding *DHFR* minor transcripts can bind to the transcription factor TFIIB, form a stable triplex with the G-track sequence in the major promoter, and impede the formation of the transcription preinitiation complex (PIC). The formation of RNA–DNA triplexes was confirmed by an electrophoresis mobility shift assay (EMSA) *in vitro* using either long RNA transcripts or short RNA oligonucleotides. The potential triplex-forming motif found in noncoding *DHFR* minor transcripts was a particularly AG-rich sequence, which suggests an antiparallel triplex formation between lncRNAs and the DNA sequence.

Promoter-associated RNAs (pRNAs) regulate the transcription of ribosomal RNA (rRNA) genes by forming stable RNA–DNA triplex with the promoter sequence (Schmitz et al., 2010). pRNAs form triple helical structures with the transcription factor binding site TTF-1 in the rRNA promoter. Most importantly, the formed RNA–DNA triplex specifically recognizes DNA methyltransferase DNMT3b, the enzyme that promotes heterochromatin formation in the rDNA promoter by DNA hypermethylation and subsequently silences the rRNA genes. The triplex-forming motif of pRNA contains mainly G, U, and C, and may form a parallel triplex with the promoter DNA based on the principle of triplex orientation.

Another interesting lncRNA, *Fendrr*, plays an important role in the development of the mammalian heart and body wall and has been suggested to regulate neighboring and distant genes by recruiting either PRC2 or the Trithorax group/Mixed lineage leukemia (TrxG/Mll) protein complexes in the target promoter (Grote et al., 2013; Grote and Herrmann, 2013). Recent studies have reported that *Fendrr* can form RNA–DNA triplexes in the promoter region of developmental genes such as *Foxf1* and *Pitx2*. The RNA–DNA triplex serves as an anchor to recruit PRC2, which in turn leaves the repressive histone marks in the abovementioned gene promoters, thereby facilitating proper tissue differentiation. Triplex formation of *Fendrr* with target DNA sequences was confirmed by *in vitro* pull-down experiments using a UC-rich triplex-forming motif. Although acidic conditions and protonation of cytosine are believed to be required for formation of triplex of UC-rich motifs, this study suggested the possibility that UC-rich motifs do form RNA–DNA triplex *in vivo*.

This possibility of RNA–DNA triplex formation by UC-rich motifs under physiological conditions was supported by a study of the antisense lncRNA *Khps1* (Postepska-Igielska et al., 2015). *Khps1* is transcribed in the antisense direction to the proto-oncogene *SPHK1* and recruits the histone acetyltransferase p300/CBP to the *SPHK1* promoter. The recruited p300/CBP acetylates the histones, which results in an active chromatin confirmation in the *SPHK1* promoter, thereby providing access to the E2F1 transcription factor, and activates *SPHK1* expression. The experimental evidence reported in that study suggested that the resultant RNA–DNA triplex could be formed by the UC-rich motif of the third-strand lncRNA *Khps1*. Interestingly, the gel shift indicating triplex formation was observed with EMSA with 10 mM Mg^{2+} at pH 7.5 without the need for an acidic condition.

Recent studies of two lncRNAs have provided considerable evidence of the formation of RNA–DNA triple helixes through AG-rich motifs of third strand lncRNAs. One study reported that the antisense lncRNA *PARTICLE* is transcribed from the promoter of *MAT2A* when cells are exposed to low-dose radiation (O'Leary et al., 2015). Surface plasmon resonance (SPR) experiments suggested that the

AG motif of *PARTICLE* forms a triple helix with the *MAT2A* promoter to recruit PRC2, which in turn methylates the CpGs of the respective promoter, thereby negatively regulating the expression of *MAT2A* after irradiation. The next lncRNA to be reported was *MEG3*, which is transcribed from human maternally expressed gene 3 (*MEG3*) (Mondal et al., 2015). In an attempt to characterize the targeting mechanism of lncRNAs in chromatin, the authors used a modified chromatin oligo affinity precipitation (ChOP) method to perform genome-wide mapping of *MEG3* binding sites; they found that these binding sites were enriched with AG-rich motifs. Subsequently, they performed various experiments using EMSA, CD spectra analysis, immunostaining with anti-triplex dA.2rU antibody and pull-down experiments, and found that AG-rich motifs guide *MEG3* to its target genes through the formation of RNA–DNA triplexes. *MEG3* can also interact with PRC2 and facilitate recruitment of PRC2 to target sites, such as the distal regulatory elements of transforming growth factor β (TGF-β) genes. The systematic studies reported by Mondal et al. strongly suggested that RNA–DNA triplex formation may be a general mechanism of target-site recognition by lncRNAs.

In addition to these direct evidence of triplex formation by lncRNAs, other reports, especially those focusing on the genome-wide mapping of target binding sites of lncRNAs, have provided indirect evidence of the possibility of triplex formation by AG-rich motifs. For example, the chromatin isolation by RNA purification (ChIRP) method has been developed for the deep sequencing of lncRNA occupancy sites in the genome (Chu et al., 2011; Simon et al., 2011). *HOTAIR* and *Drosophila roX2* lncRNAs were found to preferentially occupy AG-rich DNA motifs across the genome. Considering the number of AG-rich motifs in these lncRNAs, it is likely that they target the AG-rich DNA motifs across the genome through triplex formation. A very recent report suggested lncRNA *HOTAIR* could function on mesenchymal stem cells by forming triplex with the AG-rich motifs in the promoter of target genes (Kalwa et al., 2016). The authors predicted the triplex-forming motifs and analyzed the triplex formation by gel electrophoresis. The triplex formation by *HOTAIR* was thought to recruit PRC2 and LSD1 which are involved in gene silencing.

Although GT(U) motifs could form both parallel and antiparallel triplex with DNA duplex containing AG motifs according to the principle of triplex formation, so far there are no reports of triplex formation by lncRNAs with GU motifs. Actually, previous report excluded the possibility of triplex formation between RNA with GU motif and DNA with AG motifs by gel electrophoresis (Semerad and Maher, 1994). By testing the different combinations of RNA–DNA duplex and the third strand, the report suggested that only DNA third strand with GT motif could form stable triplex with DNA duplex with AG motif but not the RNA. These results might explain why lncRNAs with GU motifs are not favorable in the triplex formation with DNA duplex.

From these examples of RNA–DNA triplex formation by lncRNAs, one may conclude that triplex formation plays an important role in lncRNA function (Figure 2). lncRNAs inhibit the binding of transcription factors and dissociate the preinitiation complex; they also serve as guides for the recruitment of histone-modifying proteins or DNA methylation enzymes to specific sites in chromatin. For example, lncRNAs recruit PRC2 for histone modification or DNMT3b for DNA methylation in promoter regions. It has been suggested that during these epigenetic processes, triplex formation provides the general mechanism of action used by lncRNAs for the sequence-specific targeting of DNA sequences to chromatin modifiers in the genome.

Cis- and *trans*-acting mechanisms

As described above, lncRNAs can form triplexes and regulate gene expression in either a *cis*-acting manner (on neighboring genes) or a *trans*-acting manner (on distant genes) (Guttman and Rinn, 2012; Guil and Esteller, 2012). In the *cis*-acting manner, lncRNAs are usually transcribed from the promoter regions of target genes, form a triplex with promoter DNA sequences, and regulate gene expression. In the *trans*-acting manner, lncRNAs specifically target binding sites through triplex formation with distant DNA sequences. This suggests that triplex formation may be a general mechanism to explain why *trans*-acting lncRNAs can target specific genes in remote regions (Mondal et al., 2015). How the lncRNAs form triplexes with DNA sequences and whether there is any requirement for triplex formation by lncRNAs are important questions that need to be answered in order to understand the mechanism of triplex formation by both *cis*-acting and *trans*-acting lncRNAs.

With respect to the orientation of triplex formation, it is interesting that the sequence of the third strand is not the same, but rather, in reverse orientation to target DNA sequences in both the parallel and antiparallel triplex structures (Figure 1B) (Yamagata et al., 2016). For example, in the parallel triplex, the direction of the UC motif in the RNA is 5′ to 3′, whereas the direction of the TC motif in the DNA is 3′ to 5′. In the antiparallel triplex, the direction of the AG motif in the third strand RNA is 3′ to 5′, whereas the direction of the AG motif in the DNA is 5′ to 3′. This means that the third strand RNA cannot sequence-specifically target double-stranded DNA (dsDNA) of the same sequence, and it can bind only if the sequence is in the reverse orientation. This interesting observation is based on the fact that the triplex-forming motif of lncRNA can only form a triplex with the corresponding DNA sequence when it is symmetrical, because the symmetrical sequence in the reverse orientation is exactly the same as the original sequence. Triplex forming RNAs harboring such symmetrical sequence were coined as palindrome RNA, which has the ability to form intermolecular triple helix structures with DNA duplex at the transcription site (Buske et al., 2011).

The requirement for symmetrical motifs is important when considering *cis*-acting lncRNAs because they are transcribed and bind to DNA sequences in the same or neighboring regions (Figure 3A). The transcribed lncRNA with symmetrical triplex-forming motifs must change orientation and form a triplex structure with the corresponding DNA sequence. We propose that the change in orientation and symmetrical motifs are important for triplex formation by *cis*-acting lncRNAs. As shown in Table 1, the triplex-forming motifs of *cis*-acting lncRNAs are either fully or partially symmetrical. This supports our hypothesis of the need for symmetrical motifs for the formation of RNA–DNA triplex structures during gene regulation.

In the case of *trans*-acting lncRNAs, if the triplex forming motif is symmetrical, it will target dsDNA with the same sequence; if not, it will target dsDNA in the reverse direction (Figure 3B). Therefore, symmetrical motifs are also important for the identification of target sites of *trans*-acting lncRNAs. Based on the above discussion, we propose the model of the effects of symmetrical motifs on triplex formation by lncRNAs shown in Figure 3. RNA–DNA hybrid duplex formation has also been suggested as one of the potential mechanisms for lncRNA actions (Wang and Chang, 2011). The major difference between the formation of the RNA–DNA hybrid duplex and the RNA–DNA triplex is that, for the RNA–DNA hybrid duplex, the sequence and orientation are exactly the same as the corresponding DNA sequence.

Concluding remarks

Despite the large numbers of lncRNAs reported, their structures have received little attention and their structure–function relationships remain unclear (Novikova et al., 2013). Many methods have been developed for the detection and determination of RNA 2D and 3D structures, such as chemical or enzymatic probing, UV crosslinking, NMR, X-ray, and various software predictions (Mortimer et al., 2014; Zahran et al., 2015; Kubota et al., 2015; Magnus et al., 2014; Tijerina et al., 2007). However, when it comes to triplex formation by lncRNAs many important questions remain open. For example, given the potential structural complexity of lncRNAs, we are yet to understand how the RNA–DNA triplex forms and what regulates triplex formation. For example, if the triplex-forming motifs are located in the loop structure, is it possible for the loop sequence to bind directly to the DNA sequence? By contrast, if the triplex-forming motifs are involved in other secondary structures, such as hairpin structures, the structural equilibrium between RNA structures and RNA–DNA triplex formation must be considered. As shown in Table 1, some AG motifs are potentially G-rich sequences, suggesting possible competition between RNA G-quadruplex and triplex structures. Structural competition has been reported between RNA G-quadruplex and hairpin

structures, as has the fact that different ions or small molecules can regulate these preferences (Bugaut et al., 2012).

Some proteins interact directly with RNA–DNA triplex structures or regulate RNA–DNA triplex structures. For example, it has been suggested that Argonaute proteins may stabilize the miRNA–DNA triplex structure (Toscano-Garibay and Aquino-Jarquin, 2014). Some helicases have also been found to unwind the triplex formation or remodel RNA structures. It has been suggested that the helicase domain of the chromatin remodeler ATRX can remodel the structure of the repeat A region of *Xist* RNA to facilitate PRC2 binding (Sarma et al., 2014). These studies suggested the possible structural regulation by proteins in the formation or unwinding of the RNA–DNA triplex structure. The fact that RNA binds to the major groove of duplex DNA also suggests competition between RNA–DNA triplexes and protein binding. Moreover, in the case of pRNAs mentioned above, the RNA–DNA triplex structure may recruit DNMT3b and methylate the promoter CpGs, which suggests functional interactions between the triplex and proteins (Schmitz et al., 2010).

Small molecules that interact selectively with the triplex structure may also regulate or target the triplex structures formed by lncRNAs. For example, amino sugars such as neomycin are groove binders of triplex structure and other molecules with a planar structure such as quinoxaline derivatives are selective intercalators of triplex structures (Arya, 2011; Arya and Coffee, 2000; Arya et al., 2001a; Zain et al., 1999; Duval-Valentin et al., 1995; Escudé et al., 1998). Small ligands targeting miRNA triplex structures were also reported (Watkins, 2015; Nahar, 2015). These ligands might regulate the functions of lncRNAs by directly targeting the triplex structure. However, selectivity towards RNA–DNA triplex is quite important for the design of small molecules as RNA–DNA hybrid duplex, RNA triplex, DNA triplex and other hybrid triplex structures could also form *in vivo*. For example, neomycin was suggested to induce and stabilize the RNA–DNA triplex formation *in vitro* (Arya et al., 2001b). The design of selective ligands towards RNA–DNA hybrid duplex may shed light on the development of RNA–DNA triplex specific ligands (Shaw and Arya, 2008; Shaw et al., 2008). Overall, the emergence of triplex formation by lncRNAs may provide an important direction for the design of RNA–DNA triplex targeted ligands.

Accumulating evidence shows that lncRNAs may also control chromatin structure, including nucleosome positioning and chromosome looping (Böhmdorfer and Wierzbicki, 2015). The examples of triplex formation included in Table 1 also suggest that triplex formation is often associated with the recruitment of histone-modifying proteins such as PRC2. Therefore, triplex formation at the chromatin level should be considered, especially for the *trans*-acting lncRNAs that target distant genes. Previous studies have suggested that nucleosome structures can affect or even

inhibit triplex formation (Brown and Fox, 1996, 1998a, 1998b, 1999). Therefore, nucleosome and chromatin structures may affect RNA–DNA triplex formation in a manner similar to that of sequence-specific DNA binders such as polyamides (Dudouet et al., 2003; Suto et al., 2003; Gottesfeld et al., 2001; Han et al., 2014).

In summary, we have described the current evidence that supports the idea that lncRNAs triplexes exist and play significant biological roles. We also discussed the orientation of triplex formation based on the *cis*- and *trans*-acting mechanisms, and explored the possibility that proteins or small molecules can regulate triplex structures. We suggest that triplex formation represents a general mechanism for targeting DNA sequences by lncRNAs, and that triplexes may directly interact with functional proteins. However, the major concern is the need for direct evidences that support the triplex formation as the common mechanism followed by lncRNAs to target DNA sequence particularly in gene regulation. Until now, the triplex-forming motifs of lncRNAs are manually designated or predicted by software. The exact crystal or NMR structures of triplex formation in the context of lncRNA and its associated proteins are yet to be reported. Moreover, the design of selective RNA–DNA ligands that could regulate the functions of lncRNAs offers new opportunities for the development of drug targeting lncRNAs. Given the increasing numbers of lncRNAs recently reported, one may anticipate that more examples of lncRNAs involved in triplex formation will be reported and these questions could be addressed in the near future.

ACKNOWLEDGMENT

We thank Professor Peter A. Beal from Department of Chemistry, University of California (Davis) for the critical reading and valuable comments on the manuscript. This work was supported by JSPS KAKENHI (grant number 24225005).

REFERENCES

Agazie, Y.M., Burkholder, G.D., and Lee, J.S. (1996). Triplex DNA in the nucleus: direct binding of triplex-specific antibodies and their effect on transcription, replication and cell growth. Biochem. J. *316*, 461-466.

Agazie, Y.M., Lee, J.S., and Burkholder, G.D. (1994). Characterization of a new monoclonal antibody to triplex DNA and immunofluorescent staining of mammalian chromosomes. J. Biol. Chem. *269*, 7019-7023.

Arya, D. P. (2011). New approaches toward recognition of nucleic acid triple helices. Acc. Chem. Res. *44*, 134–146.

Arya, D. P., and Coffee, R. L. (2000). DNA triple helix stabilization by aminoglycoside antibiotics. Bioorg. Med. Chem. Lett. *10*, 1897–1899.

Arya, D. P., Coffee, R. L., and Charles I. (2001a). Neomycin-induced hybrid triplex formation. J. Am. Chem. Soc. *123*, 11093–11094.

Arya, D. P., Coffee, R. L., Willis, B., and Abramovitch, A. I. (2001b). Aminoglycoside-nucleic acid interactions: remarkable stabilization of DNA and RNA triple helices by neomycin. J. Am. Chem. Soc. *123*, 5385–5395.

Asensio, J.L., Brown, T., and Lane, A.N. (1999). Solution conformation of a parallel DNA triple helix with 5’ and 3’ triplex-duplex junctions. Structure *7*, 1–11.

Bacolla, A., Wang, G., and Vasquez, K.M. (2015). New Perspectives on DNA and RNA Triplexes as Effectors of Biological Activity. PLoS Genet. *11*, e1005696.

Batista, P., and Chang, H. (2013). Long Noncoding RNAs: Cellular Address Codes in Development and Disease. Cell *152*, 1298–1307.

Beal, P.A., and Dervan, P.B. (1991). Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. Science *251*, 1360–1363.

Böhmdorfer, G., and Wierzbicki, A.T. (2015). Control of Chromatin Structure by Long Noncoding

RNA. Trends Cell Biol. *25*, 623–632.

Bonasio, R., and Shiekhattar, R. (2014). Regulation of Transcription by Long Noncoding RNAs. Annu. Rev. Genet. *48*, 1–23.

Broitman, S.L., Im, D.D., and Fresco, J.R. (1987). Formation of the triple-stranded polynucleotide helix, poly(A.A.U). Proc. Natl Acad. Sci. *84*, 5120–5124.

Brosh, R.M., Majumdar, A., Desai, S., Hickson, I.D., Bohr, V.A., and Seidman, M.M. (2001). Unwinding of a DNA triple helix by the Werner and Bloom syndrome helicases. J. Biol. Chem. *276*, 3024-2030.

Brown, P., and Fox, K. (1996). Nucleosome core particles inhibit DNA triple helix formation. Biochemical Journal *319*, 607–611.

Brown, P., and Fox, K. (1998a). DNA triple-helix formation on nucleosome-bound poly(dA)·poly(dT) tracts. Biochemical Journal *333*, 259–267.

Brown, P., and Fox, K. (1998b). Triple-helix formation at different positions on nucleosomal DNA. Biochemistry *37*, 16139–16151.

Brown, P., and Fox, K. (1999). DNA triple-helix formation on nucleosome core particles. Eur. J. Biochem. *261*, 301–310.

Bugaut, A., Murat, P., and Balasubramanian, S. (2012). An RNA hairpin to G-quadruplex conformational transition. J. Am. Chem. Soc. *134*, 19953–19956.

Buske, F.A., Bauer, D.C., Mattick, J.S., and Bailey, T.L. (2012). Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. Genome Res. *22*, 1372-1381.

Buske, F.A., Mattick, J.S., and Bailey, T.L. (2011). Potential in vivo roles of nucleic acid triple-helices. RNA Biol. *8*, 427-439.

Cech, T., and Steitz, J. (2014). The noncoding RNA revolution—trashing old rules to forge new ones. Cell *157*, 77–94.

Chu, C., Qu, K., Zhong, F., Artandi, S., and Chang, H. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Molecular Cell *44*, 667–678.

Chu, C., Zhang, Q.C., da Rocha, S.T.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404–416.

Consortium, E. (2004). The ENCODE (ENCyclopedia of DNA elements) project. Science *306*, 636-640.

Consortium, E. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

De los Santos, C., Rosen, M., and Patel, D. (1989). NMR studies of DNA (R+)n.(Y-)n.(Y+)n triple helices in solution: imino and amino proton markers of T.A.T and C.G.C+ base-triple formation. Biochemistry *28*, 7282–7289.

Dixon, B.P., Lu, L., Chu, A., and Bissler, J.J. (2008). RecQ and RecG helicases have distinct roles in maintaining the stability of polypurine.polypyrimidine sequences. Mutat. Res. *643*, 20-28.

Duca, M., Vekhoff, P., Oussedik, K., Halby, L., and Arimondo, P.B. (2008). The triple helix: 50 years later, the outcome. Nucleic Acids Res. *36*, 5123–5138.

Dudouet, B., Burnett, R., Dickinson, L.A., and Wood, M.R. (2003). Accessibility of nuclear chromatin by DNA binding polyamides. Chem. Biol. *10*, 859-867.

Duval-Valentin, G., de Bizemont, T., Takasugi, M., Mergny, J.L., Bisagni, E., and Hélène, C. (1995). Triple-helix specific ligands stabilize H-DNA conformation. J. Mol. Biol. *247*, 847–858.

Escudé, C., Nguyen, C., Kukreti, S., Janin, Y., Sun, J.-S., Bisagni, E., Garestier, T., and Hélène, C. (1998). Rational design of a triple helix-specific intercalating ligand. Proc. Natl. Acad. Sci. *95*, 3591–3596.

Felsenfeld, G., Davies, D.R., and Rich, A. (1957). Formation of a three-stranded polynucleotide molecule. J. Am. Chem. Soc. *79*, 2023–2024.

Floris, R., Scaggiante, B., Manzini, G., Quadrifoglio, F., and Xodo, L.E. (1999). Effect of cations on purine.purine.pyrimidine triple helix formation in mixed-valence salt solutions. Eur. J. Biochem. *260*, 801–809.

Galupa, R., and Heard, E. (2015). X-chromosome inactivation: new insights into *cis* and *trans* regulation. Curr. Opin. Genet. Dev. *31*, 57–66.

Goñi, R., de la Cruz, X., and Orozco, M. (2004). Triplex−forming oligonucleotide target sequences in the human genome. Nucleic Acids Res. *32*, 354–360.

Gotfredsen, C.H., Schultze, P., and Feigon, J. (1998). Solution structure of an intramolecular pyrimidine-purine-pyrimidine triplex containing an RNA third strand. J. Am. Chem. Soc. *120*, 4281-4289.

Gottesfeld, J.M., Melander, C., Suto, R.K., and Raviol, H. (2001). Sequence-specific recognition of DNA in the nucleosome by pyrrole-imidazole polyamides. J. Mol. Biol. *309*, 615-629.

Grote, P., and Herrmann, B.G. (2013). The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. RNA Biol. *10*, 1579–1585.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., et al. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev. Cell *24*, 206–214.

Guil, S., and Esteller, M. (2012). Cis-acting noncoding RNAs: friends and foes. Nat. Struct. Mol. Biol. *19*, 1068–1075.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Guttman, M., and Rinn, J. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339–346.

Han, Y-W., Tsunaka, Y., Yokota, H., Matsumoto, T., Kashiwazaki, G., Morinaga, H., Hashiya, K., Bando, T., Sugiyama, H., and Harada, Y. (2014). Construction and characterization of Cy3-or

Cy5-conjugated hairpin pyrrole–imidazole polyamides binding to DNA in the nucleosome. Biomater. Sci. *2*, 297–307.

Hojland, T., Kumar, S., Babu, B.R., Umemoto, T., Albaek, N., Sharma, P.K., Nielsen, P., and Wengel, J. (2007). LNA (locked nucleic acid) and analogs as triplex-forming oligonucleotides. Org. Biomol. Chem. *5*, 2375–2379.

Hoogsteen, K. (1959). The structures of crystals containing a hydrogen complex bonding of 1-methylthymine and 9-methyladenine. Acta. Cryst. *12*, 822–823.

Hung, T., and Chang, H.Y. (2010). Long noncoding RNA in genome regulation: prospects and mechanisms. RNA Biol. *7*, 582-585.

Jain, A., Wang, G., and Vasquez, K.M. (2008). DNA triple helices: biological consequences and therapeutic potential. Biochimie *90*, 1117-1130.

Jain, A., Bacolla, A., Chakraborty, P., Grosse, F., and Vasquez, K.M. (2010). Human DHX9 helicase unwinds triple-helical DNA structures. Biochemistry *49*, 6992-6999.

Kalwa, M., Hänzelmann, S., Otto, S., Kuo, C.C., Franzen, J., Joussen, S., Fernandez-Rebollo, E., Rath, B., Koch, C., Hofmann, A., et al. (2016). The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. Nucleic Acids Res. doi: 10.1093/nar/gkw802.

Kapranov, P., Willingham, A.T., and Gingeras, T.R. (2007). Genome-wide transcription and the implications for genomic organization. Nat. Rev. Genet. *8*, 413–423.

Koshlap, K.M., Schultze, P., Brunar, H., Dervan, P.B., and Feigon, J. (1997). Solution structure of an intramolecular DNA triplex containing an N7-glycosylated guanine which mimics a protonated cytosine. Biochemistry *36*, 2659–2668.

Kowalczyk, M.S., Higgs, D.R., and Gingeras, T.R. (2012). Molecular biology: RNA discrimination. Nature *482*, 310-311.

Kubota, M., Tran, C., and Spitale, R. (2015). Progress and challenges for chemical probing of RNA structure inside living cells. Nat. Chem. Biol. *11*, 933–941.

Kung, J., Colognori, D., and Lee, J. (2013). Long noncoding RNAs: past, present, and future. Genetics *193*, 651–669.

Lacoste, J., Francois, J.C., and Helene, C. (1997). Triple helix formation with purine-rich phosphorothioate-containing oligonucleotides covalently linked to an acridine derivative. Nucleic Acids Res. *25*, 1991–1998.

Lacroix, L., Arimondo, P.B., Takasugi, M., Helene, C. and Mergny, J.L. (2000). Pyrimidine morpholino oligonucleotides form a stable triple helix in the absence of magnesium ions. Biochem. Biophys. Res. Commun. *270*, 363–369.

Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J.L., Thuong, N.T., Lhomme, J. and Helene, C. (1987). Sequence-specific recognition, photocrosslinking and cleavage of the DNA double helix by an oligo-[alpha]-thymidylate covalently linked to an azidoproflavine derivative. Nucleic Acids Res. *15*, 7749–7760.

Lee, J. (2012). Epigenetic Regulation by Long Noncoding RNAs. Science *338*, 1435–1439.

Lee, J.S., Burkholder, G.D., Latimer, L.J., Haug, B.L., and Braun, R.P. (1987). A monoclonal antibody to triplex DNA binds to eucaryotic chromosomes. Nucleic Acids Res. *15*, 1047-1061.

Li, T., Mo, X., Fu, L., Xiao, B., and Guo, J. (2016). Molecular mechanisms of long noncoding RNAs on gastric cancer. Oncotarget. *7*, 8601-8612.

Lubitz, I., Zikich, D., and Kotlyar, A. (2010). Specific high-affinity binding of thiazole orange to triplex and G-quadruplex DNA. Biochemistry *49*, 3567-3574.

Maclary, E., Hinten, M., Harris, C., and Kalantry, S. (2013). Long nonoding RNAs in the X-inactivation center. Chromosome Res. *21*, 601–614.

Magnus, M., Matelska, D., Lach, G., Chojnowski, G., Boniecki, M.J., Purta, E., Dawson, W., Dunin-Horkawicz, S., and Bujnicki, J.M. (2014). Computational modeling of RNA 3D structures, with the aid of experimental restraints. RNA Biol. *11*, 522–536.

Maine, I.P., and Kodadek, T. (1994). Efficient unwinding of triplex DNA by a DNA helicase. Biochem. Biophys. Res. Commun. *204*, 1119-1124.

Martianov, I., Ramadass, A., Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature *445*, 666–670.

Mattick, J.S. (2004). RNA regulation: a new genetics? Nat. Rev. Genet. *5*, 316-323.

Miyoshi, D., and Sugimoto, N. (2008). Molecular crowding effects on structure and stability of DNA. Biochimie. *90*, 1040-1051.

Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., Mitra, S., Mohammed, A., James, A.R., Hoberg, E., et al. (2015). MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures. Nat. Commun. *6*, 7743.

Morgan, A.R., and Wells, R.D. (1968). Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences. J. Mol. Biol. *37*, 63–80.

Morris, K., and Mattick, J. (2014). The rise of regulatory RNA. Nat. Rev. Genet. *15*, 423–437.

Mortimer, S., Kidwell, M., and Doudna, J. (2014). Insights into RNA structure and function from genome-wide studies. Nat. Rev. Genetics *15*, 469–479.

Moser, H.E., and Dervan, P.B. (1987). Sequence specific cleavage of double helical DNA by triple helix formation. Science *238*, 645–650.

Nahar, S., Ranjan, N., Ray, A., Arya, D.P., and Maiti, S. (2015). Potent inhibition of miR-27a by neomycin–bisbenzimidazole conjugates. Chem. Sci., *6*, 5837-5846.

Nakano, S., Miyoshi, D., and Sugimoto, N. (2014). Effects of molecular crowding on the structures, interactions, and functions of nucleic acids. Chem. Rev. *114*, 2733–2758.

Novikova, I., Hennelly, S., and Sanbonmatsu, K. (2013). Tackling structures of long noncoding RNAs. Int. J. Mol. Sci. *14*, 23672–23684.

O'Leary, V.B., Ovsepian, S.V., Carrascosa, L.G., Buske, F.A., Radulovic, V., Niyazi, M., Moertl, S., Trau, M., Atkinson, M.J., and Anastasov, N. (2015). PARTICLE, a triplex-forming long ncRNA,

regulates locus-specific methylation in response to low-dose irradiation. Cell Rep. *11*, 474–485.

Pasternack, L.B., Lin, S.B., Chin, T.M., Lin, W.C., Huang, D.H., and Kan, L.S. (2002). Proton NMR studies of 5'-d-(TC)(3) (CT)(3) (AG)(3)-3'—a paperclip triplex: the structural relevance of turns. Biophys. J. *82*, 3170–3180.

Ponting, C.P., Oliver, P.L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. Cell *136*, 629–641.

Postepska-Igielska, A., Giwojna, A., Gasri-Plotnitsky, L., Schmitt, N., Dold, A., Ginsberg, D., and Grummt, I. (2015). LncRNA Khps1 Regulates Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure. Mol. Cell *60*, 626–636.

Quinn, J., and Chang, H. (2016). Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. *17*, 47–62.

Radhakrishnan, I. and Patel, D.J. (1994a). Solution structure and hydration patterns of a pyrimidine.purine.pyrimidine DNA triplex containing a novel T.CG base-triple. J. Mol. Biol. *241*, 600–619.

Radhakrishnan, I., and Patel, D.J. (1994b). Solution structure of a pyrimidine.purine.pyrimidine DNA triplex containing T.AT, C+.GC and G.TA triples. Structure *2*, 17–32.

Rajendran, A., Nakano, S., and Sugimoto, N. (2010). Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. Chem. Commun. *46*, 1299-1301.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145-166.

Roberts, R.W., and Crothers, D.M. (1992). Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. Science *258*, 1463–1466.

Sarma, K., Cifuentes-Rojas, C., Ergun, A., Del Rosario, A., Jeon, Y., White, F., Sadreyev, R., and Lee, J.T. (2014). ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. Cell *159*, 869–883.

Schmitz, K.M., Mayer, C., and Postepska, A. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. *24*, 2264-2269.

Semerad, C.L., and Maher, L.J. (1994). Exclusion of RNA strands from a purine motif triple helix. Nucleic Acids Res. *22*, 5321-5325.

Shaw, N.N., and Arya, D.P. (2008). Recognition of the unique structure of DNA:RNA hybrids. Biochimie *90*, 1026-1039.

Shaw, N.N., Xi, H., and Arya, D.P. (2008). Molecular recognition of a DNA:RNA hybrid: sub-nanomolar binding by a neomycin-methidium conjugate. Bioorg. Med. Chem. Lett. *18*, 4142-4145.

Shimizu, M., Konishi, A., Shimada, Y., Inoue, H., and Ohtsuka, E. (1992). Oligo (2'-O-methyl) ribonucleotides: Effective probes for duplex DNA. FEBS Lett. *302*, 155–158.

Simon, M., Wang, C., Kharchenko, P., West, J., Chapman, B., Alekseyenko, A., Borowsky, M., Kuroda, M., and Kingston, R. (2011). The genomic binding sites of a noncoding RNA. Proc. Natl. Acad. Sci. *108*, 20497–20502.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat. Struct. Mol. Biol. *14*, 103-105.

Suto, R.K., Edayathumangalam, R.S., and White, C.L. (2003). Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. J. Mol. Biol. *326*, 371-380.

Takahama, K., and Oyoshi, T. (2013). Specific binding of modified RGG domain in TLS/FUS to G-quadruplex RNA: tyrosines in RGG domain recognize 2'-OH of the riboses of loops in G-quadruplex. J. Am. Chem. Soc. *135*, 18016–18019.

Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R., and Oyoshi, T. (2013). Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. Chem. Biol. *20*, 341–350.

Tijerina, P., Mohr, S., and Russell, R. (2007). DMS footprinting of structured RNAs and RNA-protein complexes. Nat. Protoc. *2*, 2608–2623.

Torigoe, H., Hari, Y., Sekiguchi, M., Obika, S., and Imanishi, T. (2001). 2'-O,4'-C-methylene bridged nucleic acid modification promotes pyrimidine motif triplex DNA formation at physiological pH: thermodynamic and kinetic studies. J. Biol. Chem. *276*, 2354–2360.

Toscano-Garibay, J.D., and Aquino-Jarquin, G. (2014). Transcriptional regulation mechanism mediated by miRNA–DNA• DNA triplex structure stabilized by Argonaute. Biochim. Biophys. Acta. *1839*, 1079-1083.

Tsai, M.-C.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689–693.

Van Bemmel, J.G., Mira-Bontenbal, H., and Gribnau, J. (2016). Cis- and trans-regulation in X inactivation. Chromosoma. *125*, 41–50.

Wang, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. Mol. Cell *43*, 904–914.

Watkins, D., Jiang, L.W., Nahar, S., Maiti, S., and Arya, D.P. (2015). A pH sensitive high-throughput assay for miRNA binding of a peptide-aminoglycoside (PA) library. PLoS One *10*, e0144251.

Watson, J.D., and Crick, F. (1953). Molecular structure of nucleic acids. Nature *171*, 737-738.

Wells, R.D., Collier, D.A., Hanvey, J.C., Shimizu, M., and Wohlrab, F. (1988). The chemistry and biology of unusual DNA structures adopted by oligopurine oligopyrimidine sequences. FASEB J. *2*, 2939–2949.

Wilusz, J.E., Sunwoo, H., and Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. *23*, 1494–1504.

Yamagata, Y., Emura, T., and Hidaka, K. (2016). Triple helix formation in a topologically controlled DNA nanosystem. Chemistry *22*, 5494-5498.

Zahran, M., Sevim Bayrak, C., Elmetwaly, S., and Schlick, T. (2015). RAG-3D: a search tool for RNA 3D substructures. Nucleic Acids Res. *43*, 9474–9488.

Zain, R., Marchand, C., Sun, J., Nguyen, C. H., Bisagni, E., Garestier, T., and Hélène, C. (1999). Design of a triple-helix-specific cleaving reagent. Chem. Biol. *6*, 771–777.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750-756.

FIGURE LEGENDS

Figure 1. Triplex formation between the third strand oligonucleotide and dsDNA.

(A) Examples of Hoogsteen and reverse Hoogsteen hydrogen bonding in the triplex formation.

(B) Illustration of parallel and antiparallel triplex formation. The arrow indicates 5’ to 3’ direction.

(C) Structure of RNA–DNA triplex. The third strand RNA (brown) binds to the major groove of duplex DNA. Picture was generated by the reported RNA-DNA triplex structure (PDB 1R3X) (Gotfredsen et al., 1998).

Figure 2. Examples of lncRNAs that are reported to form RNA–DNA triplexes.

lncRNAs can form a triplex structure with a DNA sequence in either neighboring or distant genes and regulate their expression. lncRNAs also interact with various regulatory proteins such as transcription factors and chromatin modifiers.

Figure 3. Orientation of triplex formation by lncRNAs.

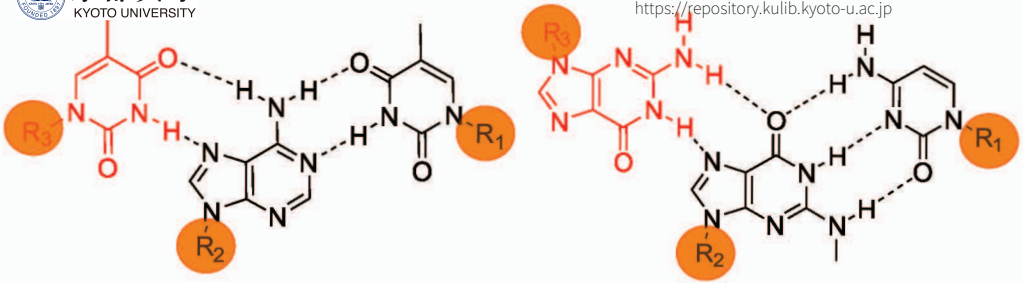
(A) If a transcribed lncRNA forms a triplex with a DNA sequence in the same region, it must change its orientation; in this case, the symmetrical triplex-forming motif is favorable.

(B) In the case of the *trans*-acting mechanism, a change in orientation is also needed for triplex formation. The lncRNA can form a triplex with the same symmetrical sequence or a reversed asymmetrical sequence.

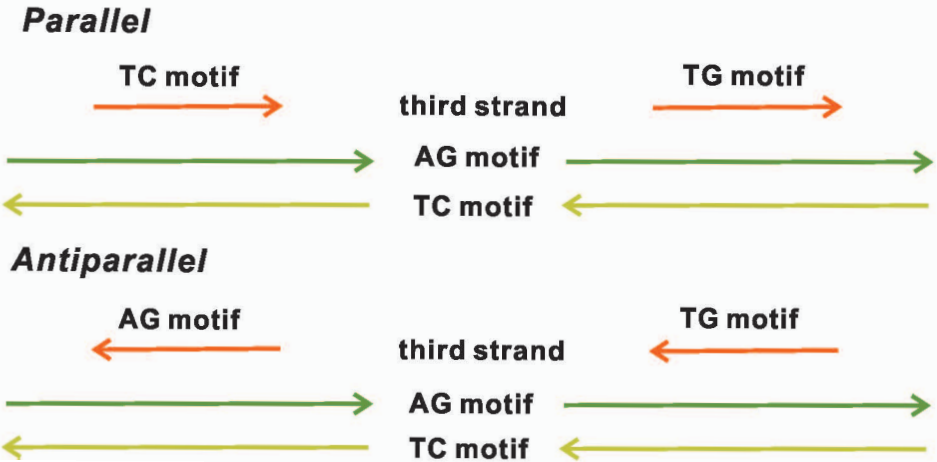
Table 1. Examples of triplex formation by lncRNAs.

lncRNA	**Triplex forming motifs (5' to 3')**	***cis* vs *trans***	**Associated proteins**
pRNA of *DHFR*	ACAAAUGGGGACG AGGGGGGCGGGGC GGCC	*cis* on *DHFR* gene	Interact with TFIIB and dissociate the preinitiation complex (Martianov et al., 2007)
pRNA of rRNA gene	GUCGACCAGUUGU UCCUUUG	*cis* on rRNA gene	Recruit DNMT3b for DNA methylation (Schmitz et al., 2010)
Fendrr	UCCCCUCCAUCCUC UUCCUUCUCCUCCU CCUCUUCUUU	*cis* on *Foxf1* gene; *trans* on *Pitx2* gene.	Bind to PRC2 and TrxG/MLL (Grote et al., 2013)
Khps1	UCCCCCUUUUUUU UUCCUCCU	*cis* on *SPHK1* gene	Recruit p300/CBP to the *SPHK1* promoter (Postepska-Igielska et al., 2015)
PARTICLE	AAGGGGGGGGGAA	*cis* on *MAT2A* gene	Interact with G9a and SUZ12 (subunit of PRC2) (O'Leary et al., 2015)
MEG3	CGGAGAGCAGAGA GGGAGCG	*trans* on TGF-β pathway genes	Recruit the PRC2 to target site (Mondal et al., 2015)
HOTAIR	AGAGGAGGGAAGA GAG	*trans* on *PCDH7, HOXB2* et al.	Bind to PRC2 and LSD1 (Kalwa et al., 2016)

A



B



C

Figure 1

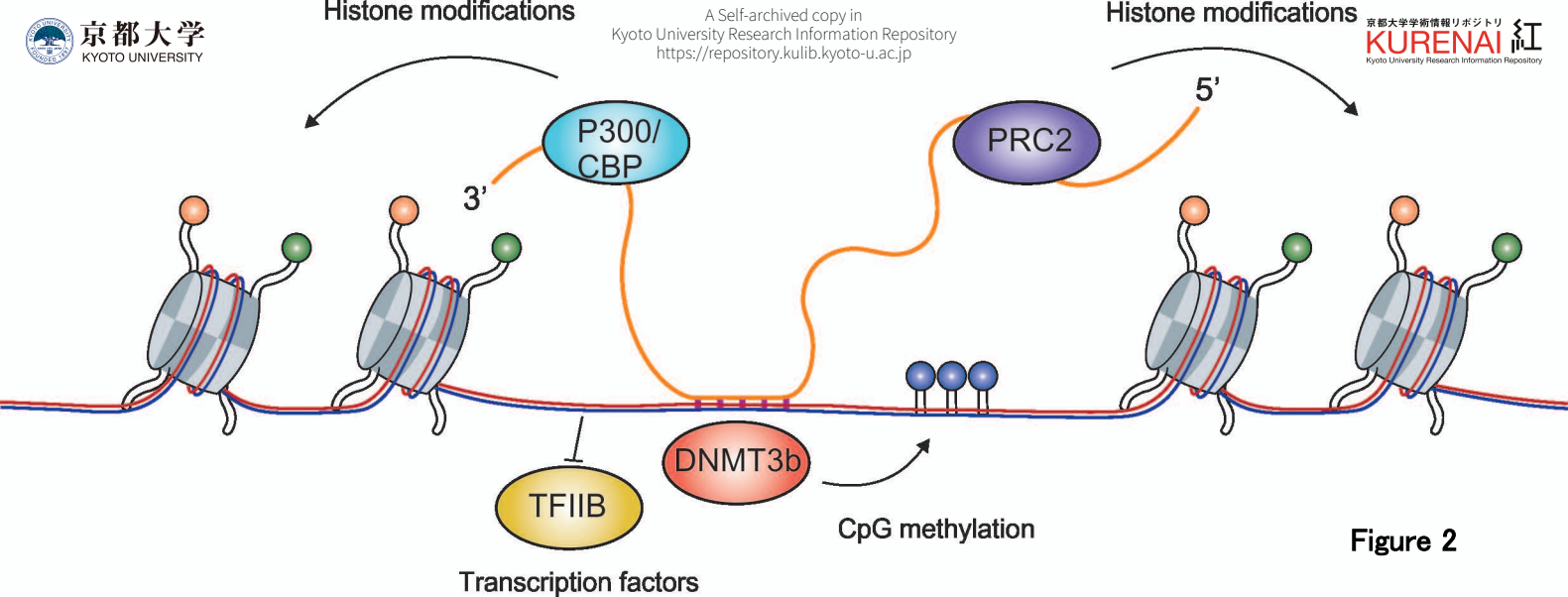
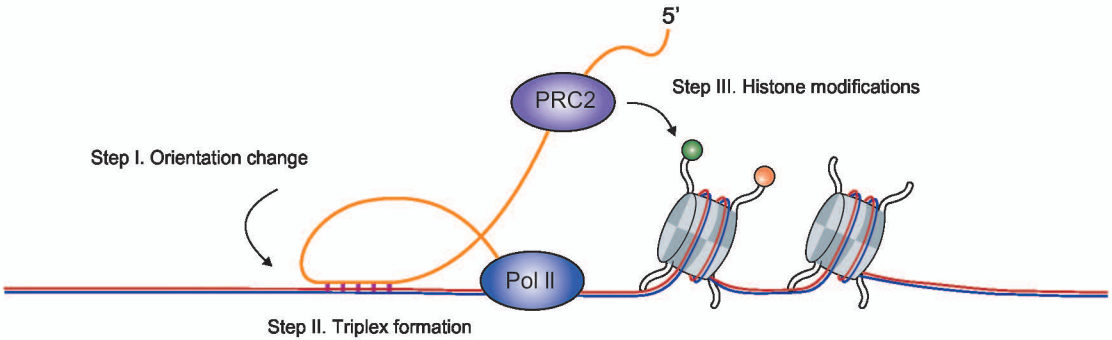


Figure 2



B

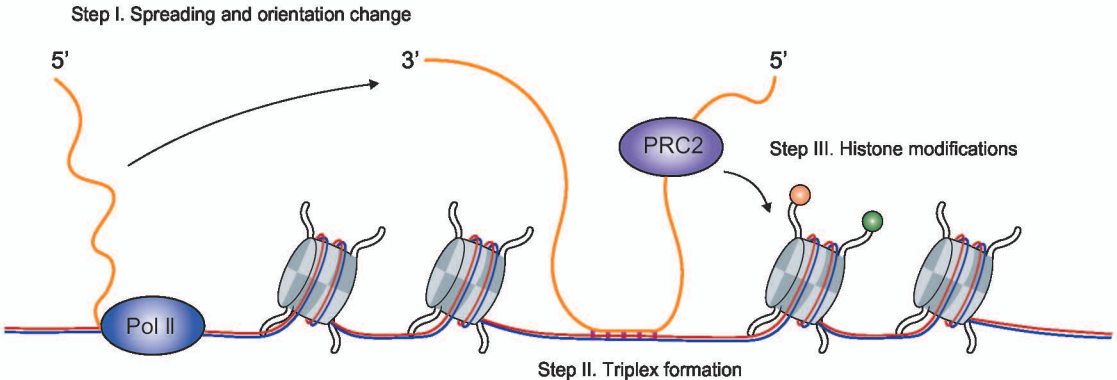


Figure 3